DOCUMENT RESUME

ED 057 698 FL 002 812

AUTHOR Carroll, John B.
TITLE Defining Language Comprehension: Some Speculations.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO RM-71-9
PUB DATE Jul 71
NOTE 55p.; Paper presented at the Research Workshop on Language Comprehension and the Acquisition of Knowledge, Quail Roost Conference Center, Durham, North Carolina, March 31-April 3, 1971

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cognitive Processes; Comprehension Development; Intellectual Development; Language Development; Language Proficiency; *Language Research; *Learning Processes; Learning Theories; Linguistic Competence; *Listening Comprehension; Perception; *Reading Comprehension; Second Language Learning

ABSTRACT

This is a survey of techniques that have been used to test language comprehension. The study of research completed in this field points up the fact that there is no single technique that universally gives valid and reliable information. Various definitions of language comprehension are examined with special emphasis placed on implications for the teacher and the learner. The author develops a classification of procedures for testing comprehension on the basis of a survey of procedures followed in psychometric and experimental investigations. (RL)

ED057698

RM-71-9

RESEARCH MEMORANDUM

DEFINING LANGUAGE COMPREHENSION: SOME SPECULATIONS

John B. Carroll

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

This paper was presented at the Research Workshop on Language Comprehension and the Acquisition of Knowledge, Quail Roost Conference Center, Durham, North Carolina, March 31-April 3, 1971

FL002812

Educational Testing Service
Princeton, New Jersey
July 1971

Defining Language Comprehension: Some Speculations

John B. Carroll
Educational Testing Service

The concept of comprehension is of major relevance to education. In the most general sense of "being educated," an "educated" person possesses a certain body of knowledge, competences, abilities, and skills. On the one hand, this implies some sort of structure that has been laid down in the individual, presumably in his nervous system, or, one might say, in a memory store, as a result of his whole prior development and experience, including educational experiences. Let us assume that this structure includes, among other things, a "cognitive structure" that consists of a large number of "comprehensions" or "understandings" of the almost infinitely diverse phenomena to which the individual has been, or is likely to be exposed. In the study of comprehension processes we must take account of the nature of this structure--noting, however, that it is with the structure of the individual's knowledge that we are concerned, not the "structure of knowledge" in general, for that is an abstraction that may or may not have any isomorphism with the individual's cognitive structure. On the other hand, "being educated" implies a capacity for acquiring new understandings and integrating them in some valid way with the knowledge already acquired. One aspect of this capacity is certainly the ability to understand language (normally, at least the native language, but other languages may be included in the individual's repertoire), and through that ability to acquire new knowledge. It is

with this language comprehension process, and the process of acquiring knowledge through language, that this conference is concerned. We recognize, of course, that there are other modes of acquiring knowledge, but we limit ourselves to the consideration of comprehension through language except to the extent that such comprehension is supported, facilitated, or otherwise affected by these other modes of apprehending.

Educators have long wrestled with the problem of language comprehension. They have recognized that the child's competence in his native language, at the time of school entrance, is far from sufficient to permit him to acquire, through language, the range and complexity of knowledge and skills that are contained in the total school program. Consequently, a major concern of the school curriculum is with the promotion of what are essentially language comprehension skills at progressively higher levels of grammatical, lexical, and semantic knowledge. Beyond the process of teaching the child to decode print into some analogue of spoken language, educators find that there still remains the problem of teaching the child to "understand" the language thus decoded. "Listening comprehension" and "reading comprehension" are two phrases that appear very frequently in educational literature, but there is much study and debate as to what those phrases might mean. Their definition becomes particularly problematical when one attempts to develop measures of listening comprehension or of reading comprehension. Davis (1941) was able to assemble a list of several hundred "reading comprehension skills," but since many of these overlapped, he grouped them into nine "testable skills," and in a factor analytic study (Davis, 1944) he felt he had confirmed the independent existence of these nine skills. Using a different factor-analytic approach, Thurstone (1946) claimed that these nine skills represented only

one, or at most two independent factors of reading ability. In subsequent work, Davis (1968) reaffirmed the independent existence of eight of these skills, but if one considers the amount of unique variance residing in the tests of these skills one is tempted to conclude that perhaps only four or five of them merit recognition as distinct skills, and even these are rather highly correlated in high-school populations. These "factors" are: "remembering word meanings," "following the structure of a passage," "finding answers to questions answered explicitly or in paraphrase," "recognizing a writer's purpose, attitude, tone and mood," and "drawing inferences from the content."

The story is roughly the same in the field of "listening comprehension" testing. In planning the development of the so-called STEP Tests of Listening published by ETS (1956-59), a committee drew up an impressive list of "listening comprehension skills" that were to be represented in these tests, skills such as "plain-sense comprehension" (identifying main ideas, remembering details and simple sequences of ideas, understanding word meanings); "interpretation" (understanding implications of main ideas and significant details, interrelationships among ideas, and connotative meanings of words); and "evaluation and application" (judging validity of ideas, distinguishing fact from fancy, noting contradictions, judging whether the speaker has created the intended mood or effect, etc.). It can be seen that this is a true hodge-podge, but in view of the fact that the test committee had no real theory of listening comprehension on which to draw, this is pardonable. Other listening comprehension tests have been devised, such as the Brown-Carlsen test (Brown & Carlsen, 1953); what is rather disturbing, however, is that the various tests of "listening ability" tend to show no higher intercorrelations among themselves than they show with reading and intelligence tests (Kelly, 1965). The evidence suggests that

listening tests measure a mixed bag of functions (Bateman, Frandsen, & Dedmon, 1964; Freshley & Anderson, 1968), but are mainly measures of "verbal ability."

In this connection it is necessary to point out that tests of listening comprehension and reading comprehension are designed to measure generalized skills of comprehension ability. The test maker is not concerned with measuring how well the examinee comprehends a particular spoken or written text; rather, he is concerned with the examinee's ability to comprehend a sample of such texts, in order to infer the examinee's ability to understand additional texts. Measuring comprehension ability is in some respects a problem quite different from that of measuring the degree of comprehension that a subject has when exposed to a given language stimulus. This latter problem will be considered in another section of this paper. But with regard to ability measurements, it should be mentioned that most presently available tests do not permit a satisfactory assessment of the individual's "absolute" level of comprehension ability. Even if it is assumed that comprehension ability is a unitary dimension of individual differences, tests do not permit the placement of an individual on a scale that would indicate in meaningful terms, for example, the difficulty level of textual materials that the individual would be able to comprehend to some desired criterion. The lack of such tests has made it difficult to assess accurately the distribution of levels of "literacy" in the U.S. population at different age levels.

Comprehension ability, however, is more likely a multidimensional affair. Whether one is concerned with spoken or printed language, the evidence suggests that the individual may have different levels of

ability with respect to vocabulary, grammatical features, and other characteristics of texts. In listening comprehension, attentional, motivational, auditory, and memory factors may be involved (Spearritt, 1962). In reading comprehension, speed and level of comprehension have long been recognized as conceptually distinct even if they are not statistically independent (Blommers & Lindquist, 1944). Comprehension ability tests tend to be substantially correlated with "intelligence" tests, even those of a nonverbal character, such as a figure analogies test. This is not the place to try to interpret such a finding in depth. However, it is à propos to mention that one possible source of this correlation is the fact that reading and listening comprehension tests do not measure only what may be called "pure" comprehension of language; because of the way in which they are constructed, and the kind of items they include, they tend also to measure ability to make inferences and deductions from text content. A question that this conference should address is whether it is possible in fact to distinguish "pure" comprehension of language texts from processes of inference, deduction, and problem solving that often accompany the reception of language. An empirical research question would be to see whether it would be possible to decrease the correlation of comprehension ability tests with intelligence tests by eliminating or reducing those elements of comprehension tests that call for inferential processes that go beyond sheer comprehension. This problem has not, to my knowledge, been investigated.

Depending on the method of their administration, comprehension ability tests may also involve memory abilities. Research is needed to see to what extent it is possible to reduce their dependence on memory.

An adequate theory of language comprehension would undoubtedly be of help in the construction of comprehension ability tests. Bormuth (1970) has attempted to develop a systematic theory for this purpose. His approach utilizes the theory of transformational-generative grammar. In essence, he recommends that if one is interested in testing comprehension of a sentence or a longer discourse (or, indeed, a complete course of instruction in a subject-matter), the test questions should be based on transformations of sentences in the text to which the student has been exposed. For example, given the base sentence (1):

(1) A very old man who lives up the street led his dog up to a store window one day.

one could form, through systematic applications of transformation rules, such questions as (1a - 1c):

(1a) Who led his dog?

(1b) What did the man lead?

(1c) Where does the man live?

etc.

Thus far Bormuth has offered only very simple examples of his technique, employing relatively simple grammatical transformations. One might suppose that such simple transformations would be within the reach of almost any native speaker beyond the stage of primary language acquisition. Nevertheless, in a study of fourth-grade children's ability to understand various syntactic structures, using these techniques, Bormuth, Manning, Carr, and Pearson (1970) concluded that "large proportions of the children were unable to demonstrate a comprehension of even these basic structures by which information is signaled...." I suspect, however, that much more elaborate transformations, probably of a

"semantic" character, would be required to provide effective comprehension test questions at higher levels of ability. Further development of Bormuth's approach would undoubtedly require a considerable amount of special-purpose linguistic research, as well as research in the psychometric application of the results.

Another important educational problem for which a theory of language comprehension might be able to give solutions is the problem that is referred to by the phrase "mere verbalization." By this is meant a kind of learning that goes only so far as to observe the words, and not the meaningful content, of didactic discourse. It is commonly noted that children can memorize rules and definitions without any evidence of true comprehension of them or of ability to apply them properly. How should we interpret this phenomenon? Is it simply another case of deficient language comprehension competence, is it a function of "set" or motivation, or is it a case of poor performance, i.e., errors in the application of knowledge?

This leads us to the more general problem of how we understand language and what we mean when we say we derive knowledge from language. Obviously this problem pervades education at all levels, because in view of the way in which educational programs are conducted, with lectures, readings, film narrations, and manifold other uses of language, it must be the case that educators have high expectations as to the efficacy of language communications. Yet it is obvious that learning from language does not always occur efficaciously. How shall we analyze these failures? To what extent are they due to deficits in language competence, and to what extent are they due to performance factors, the conditions of

instruction, etc.? Questions such as these, it seems to me, are within the purview of this conference.

The Problem of Defining Language Comprehension

In approaching the definition of language comprehension, we may start with the observation that a mature language user can and often does render a judgment as to whether he does or does not comprehend a particular stretch of discourse. He may render this judgment with respect to a particular word, a phrase, a clause, a whole sentence, or a longer discourse. If a reader fails to understand a particular word, perhaps he will go and look it up in a dictionary or other reference work. Failure to understand a phrase or some longer stretch of discourse may prompt the reader to reinspect the preceding context, exhibiting "regressive" eye movements. In the case of a hearer, failure to understand something may prompt him to request clarification from the speaker (if present and available). Such behaviors are at least evidence for the proposition that an attentive language receiver continually monitors his own comprehension processes and is generally aware of whether he "comprehends" or not. It is also evidence that suggests that comprehension is an internal, subjective process that is in general not open to external observation. Even the detection of subvocal speech movements during silent reading by electromyography (Edfeldt, 1960; McGuigan, Keller, & Stanton, 1964) is only a very indirect and unreliable method of indexing comprehension.

At this stage of the discussion I am not claiming that the language receiver's judgment is veridical. At any point he may be misunderstanding the intent of the discourse even though he believes himself to be

comprehending (the false positive case), and it is even possible that he actually understands even though he believes himself not to understand (the false negative case). Nevertheless, let us assume that in most cases the language receiver's judgments are reliable and veridical.

The simplest possible test of comprehension, therefore, is to have the language receiver render his subjective judgments of comprehension in an overt manner. This idea has been applied in certain kinds of experimental settings. For example, in unpublished work on "comprehension tracking" done by Daniel Forsyth and Herbert Rubenstein at the Harvard Center for Cognitive Studies (see the Center's 7th Annual Report, 1966-67, pp. 26-27) sentences are presented one, two, and four words at a time by means of a computer-controlled CRT display. The subject observes the display and presses a button as soon as he thinks he comprehends it, causing the next segment to appear. The time that each segment is displayed, i.e., the time taken by S to report comprehension, is recorded by the computer and these times can be related to characteristics of the sentence fragments that have been presented—their length, their position in the sentence, their grammatical characteristics, etc. Danks (1969) presented subjects with short printed sentences and measured "comprehension time" by asking them to press a key as soon as they comprehended a given sentence. Some of the sentences were grammatically well-formed, meaningful sentences; others were deviant with respect to either grammar or meaning, or both. Danks found that the latencies for sentence comprehension were primarily a function of their meaningfulness; grammaticalness was only of secondary importance. He insured that the Ss kept "honest" in their reports of comprehension by requiring them to paraphrase the sentences on 40% of the trials. It is interesting, incidentally, that

Ss reported "comprehension" even of presumably meaningless, ungrammatical sentences such as "Guests tall fair sail goats." They did this either by misperceiving words (e.g., mistaking goats for boats) or by conjuring up highly fanciful interpretations (e.g., "Tall fair guests sail ships in the shape of goats"). This suggests that comprehension contains an element of problem solving.

There are obvious difficulties with subjective reports, even when accompanied by test probes, latency measurements, and the like. It would be inappropriate to use subjective reports in an adversary testing situation: imagine the chaos that would result if ETS asked students taking the SAT simply to report how well they understood reading comprehension paragraphs! Therefore we will want to consider more objective methods of testing comprehension.

Before doing so, perhaps we should make a preliminary characterization of language comprehension so that we may have some idea of what we are after in attempting to select more objective techniques of testing. It is particularly important to identify what accompanying processes we may wish not to test or measure. I can think of two candidates for such processes: memory and inference.

Memory. If comprehension is a process that occurs more or less simultaneously with the reception of a message, we would be interested in the occurrence or nonoccurrence of that process only during the reception of the message or at least within a very short time-lag. Thus, if memory is to be involved at all, it should be only what has been called short-term memory, i.e., memory that can fade within a few seconds. As soon as longer time-intervals are involved in the testing of comprehension, there is the possibility that we are studying memory processes along with, or

in place of, comprehension processes. For example, it is conceivable that there could be completely satisfactory comprehension at the time of message reception, but complete or nearly complete loss of that comprehension after the fading of short-term memory.

Some of the methodological problems in the use of memorial techniques to assess the comprehension of syntactic structures have been elucidated by Fillenbaum (1970). He shows, for example, that affirmative and negative yes/no questions are actually understood in different ways even though they appear to be similar in certain studies employing memory techniques. One may also be reminded of Epstein's (1969) experiment that suggested that the Savin and Perchonock (1965) "effect," whereby different types of sentences are claimed to occupy space in memory storage as a function of their transformational complexity, reflects retrieval rather than storage and comprehension processes.

There is also the possibility that there could be memories without comprehension, whatever comprehension may turn out to be. Marks and Jack (1952) give some data concerning immediate memory span for strings of various orders of "approximation to English," and although memory span increases with order of approximation, the results can be interpreted as suggesting that even when a sentence is not comprehended, rendition of at least a part of that sentence in immediate memory span can take place on the basis of pure memory. It is well known that with rehearsal and multiple trials, subjects can learn to reproduce much longer passages <u>verbatim</u> and without comprehension, e.g., materials in a foreign language. It is curious, however, that, according to King and Russell (1966, p. 482), Ss instructed to learn connected meaningful material for its substance and ideas "tend to recall proportionately more words, letters, sentences,

etc., than ideas or sequences of words," whereas Ss instructed to learn verbatim "recall proportionately fewer words, letters, sentences, etc., and more ideas."

Nevertheless, it is possible to take an entirely opposite view on the question of whether memory factors should be included in tests of comprehension. It can be argued that, at least in educational contexts, there is little use in comprehending a message unless the outcome of that comprehension is remembered and transferred to a long-term memory store. Certainly the evidence from a large number of studies employing memorial techniques is to the effect that material that is more "meaningful" and hence more easily comprehended is more likely to be retained. Thus, comprehension appears to facilitate memory even though it may be neither necessary nor sufficient for memory to occur.

Moreover, there is evidence to the effect that what is remembered from exposure to connected discourse tends to be its "meaning" content rather than the particular phraseology in which that meaning is couched. The work of Bartlett (1932), Gomulicki (1956), and Paul (1959), among others, shows that both in storage and retrieval processes subjects who are asked to learn connected discourse operate much more with "ideas" and basic meanings than with the verbatim phraseology. Sachs (1967a, 1967b) has shown that memory for syntactic and specific lexical content in prose fades very rapidly even when tested by recognition techniques, whereas memory for meaning persists much longer. What all this suggests is that the study of comprehension as such may profit from the judicious use of memorial techniques; with appropriate control of temporal factors one may largely eliminate the effect of quite superficial features of discourse, i.e., its

surface structure in grammar and lexis, freeing one to deal only with deeper aspects of meaning. (Whether these deeper aspects of meaning are actually equivalent to the "deep structure" of transformational grammar is a question that I will not try to open at this point.) This conclusion actually has minimal conflict with the recommendations of Fillenbaum (1970) cited earlier, because Fillenbaum was concerned with the assessment of the understanding of syntactic features whose meaning components are relatively superficial, such as the difference between the sentences "Is the shop closed?" and "Isn't the shop closed?" that merely signals the speaker's expectation as to the answer.

Even though this discussion started with an argument against the use of memory techniques, we come out with a less trenchant attitude. On balance, we have to realize that memory factors can hardly be avoided, even when we try to restrict the testing of comprehension to an "immediate" test. For example, suppose we construct a typical reading comprehension test with paragraph stimuli and multiple-choice questions over the paragraphs. The test questions could be administered either with or without allowing the examinee to reexamine the paragraphs after he has had his initial opportunity to read and study them. If we do <u>not</u> permit reinspection of the paragraphs, we would certainly be emphasizing memory factors. The more typical manner of administering a reading comprehension test, however, is to allow inspection of the paragraphs along with the questions. Even this method does not completely eliminate memory because the examinee may still have to remember where in the paragraphs to look for a desired answer, and there is even the possibility of memory loss

between the act of finding an answer and utilizing it in answering a question. Note that in the case of listening comprehension tests it is rarely possible for the examinee to rehear the initial material as he answers questions; in measuring listening comprehension we are virtually forced to allow memory factors to operate. Comparisons between reading and listening comprehension tests would have to control this factor.

Inference and related reasoning processes. I said above that we might want to consider eliminating inference and related reasoning processes from tests of comprehension. I had earlier suggested that many reading and listening comprehension ability tests may be for some purposes too heavily loaded with demands on the individual's reasoning processes, so that they tend to measure general verbal intelligence and reasoning skills rather than comprehension per se. Of course, it is possible that with the elimination of reasoning processes there would be nothing left, but I tend to doubt this in view of the factor analytic studies (e.g., Carroll, 1941) that have clearly separated inductive and deductive factors from "verbal ability." I would also appeal to the work of Davis (1968), who, at least according to my interpretation (Carroll, 1969), was able to separate several "pure" comprehension factors (depending, respectively, on lexical knowledge, grammatical knowledge, and an ability to "locate facts" in paragraphs) from an inferential factor requiring the examinee to go beyond the data given.

The problem of whether one wants to include "inference" in comprehension may be presented in a relatively simple form when we consider the three-term inference problem studied by Clark (1969), among others. That is, if we present a sentence like (2):

(2) John isn't as tall as Mary, but he is taller than Tom.

and then pose a question such as "Who is tallest?" or "Who is shortest?" or "Who is in-between?", producing the answer seems to require more than a simple "parsing" of the sentence. That is, a subject might fully "comprehend" the meanings of the two clauses without doing the additional processing of information required to answer such questions. The additional processing, perhaps, is dependent upon the question asked. Suppose one simply asked, "Who is shorter than Mary?" It seems likely (though I don't believe this experiment has been done) that the readiest answer would be "John," based solely on the first clause, though "Tom" or "both John and Tom" would also be acceptable answers. Yet, even the processing of the first clause to yield the answer "John" intuitively requires a certain amount of intellectual effort that again goes beyond sheer comprehension, more effort, let us say, than answering the question, "Is John taller than Mary?" Clark's data suggest that there is a continuum ranging from comprehension of the simple surface structure in terms of what he calls its "functional relations" up through inferential processes of considerable complexity, whose stages can be identified by experimental techniques. (I am sure we will hear more about this from Trabasso.) The problem we face is whether it is actually useful to draw a line between what I have called "simple comprehension," on the one hand, and "inferential processes," on the other, and if so, where on the continuum the line should be drawn. But even the three-term inference problem studied by Clark is by no means the most involved kind of inference required in standard reading comprehension tests. Consider the following

item offered by Davis (1968) as measuring the skill of "making inferences about the content":

> The delight Tad had felt during his long hours in the glen faded as he drew near the cabin. The sun was nearly gone and Tad's father was at the woodpile. He was wearing the broadcloth suit that he wore to church and to town sometimes. Tad saw his father's hands close around a bundle of wood. He was doing Tad's work--and in his good clothes. Tad ran to him. "I'll git it, Pa."
>
> When Tad saw his father, he felt
> A disappointed
> B impatient
> C angry
> D guilty

It would seem extremely difficult (although conceivably it could be done) to specify any linguistic rules whereby the "correct" answer to this item could be predicted from the paragraph. Selecting the most likely correct answer seems to require, on the part of a test subject, not merely a literal comprehension of the paragraph and the question but also an apprehension of the total situation described in the paragraph and a sensitivity to social relationships and expectations that are only hinted at in the paragraph. (In fact, the keyed answer, "guilty," is not the only answer that might conceivably be correct, given the statements in the paragraph. If Tad's father were a drunkard habitually given to acting on impulse and if Tad had promised his father that he would do his chores even if he were late, he might feel impatient, angry, or disappointed rather than guilty. This consideration adds weight to the assertion that an example of this sort suggests that inferential processing of information requires much more than literal comprehension.)

At least two important points emerge from this digression to explore processes that might accompany language comprehension:

(1) Language comprehension occurs in situational contexts whose characteristics may influence not only the degree to which comprehension processes operate but also the nature and extent of certain other processes that may accompany comprehension, usually as a consequence of it. The special arrangements that are frequently necessary to test comprehension constitute such situational contexts.

(2) Two processes often co-occurring with comprehension are memory and inference; while they are conceptually distinguishable from comprehension, their occurrence may make it difficult to assess the separate occurrence of the comprehension process itself.

Let us now address ourselves to attempting to make a preliminary characterization of language comprehension itself. I shall not attempt, however, to analyze the comprehension process, i.e., to specify how the individual arrives at a state of comprehension. This is a problem that has received much discussion, for example, in various papers presented at the Edinburgh University Conference on Psycholinguistics (Lyons & Wales, 1966), and it will undoubtedly be the concern of some of the other papers to be presented here. For the purpose of providing a framework for assessing tests of comprehension, I am only interested in characterizing the end state of the comprehension process, that is, in specifying what the individual can be expected to have accomplished in comprehending a particular stretch of discourse.

To make the task somewhat less complicated than it might otherwise be, let us assume initially that the message is both "meaningful" and grammatically well-formed. Later we will consider cases in which there may be deviation from full meaningfulness and grammatical well-formedness.

The commonly accepted definition of comprehension is that it is the process of apprehending the "meaning" of something--the "meaning" of a word, of a phrase or idiom, of a sentence, or of a longer discourse. This implies that in order to assess the comprehension of a given segment of a verbal message, we must identify the "meaning" that is to be comprehended. The identification of meaning is a difficult and tangled problem, but I see no alternative to trying once more to explicate what is meant by meaning in the case of verbal discourse, at least to the extent of having a workable concept for use in assessing procedures for testing comprehension.

Discussions of meaning have often been encumbered by a failure to distinguish between the meaning of a given linguistic element that is implicit in the rules of its use in the speech-community and the total meaning of a discourse (of whatever length) composed of such elements. The kind of distinction I have in mind was referred to by Miller (1965, p. 18) when he urged that "the meaning of an utterance is not a linear sum of the meanings of the words that comprise it," but I feel that these different meanings of <u>meaning</u> need further explication.

First consider the "meaning of a given linguistic element." By "linguistic element" I mean any linguistic unit that has a meaning in the sense that one or more rules or conventions can be specified as to the relation of that unit with a concept or class of experiences as developed by members of the speech-community. The meaning of the linguistic unit would be incorporated in these rules or conventions. I do not wish to commit myself to any particular linguistic theory in saying this, nor to prompt a discussion of linguistic theories and techniques. I simply assume that however one analyzes a linguistic

system, there are going to be certain units or elements whose correspondence with classes of speaker experiences can in theory be specified; examples of units might include, for example, what structural linguists have called morphemes and grammatical constructions, or what transformational linguists call formatives, base structures, etc., with meanings that could be quite concrete or quite abstract. A part of the "competence" of the language user is the "knowledge" of a large collection of these rules relating form and meaning. (I shall not try to specify how this "knowledge" should be characterized in psychological terms; it is not relevant here to discuss whether it is best conceptualized in terms of "cognitive structure," "habit," "response disposition," or whatever else might be proposed.)

We cannot, of course, expect every language user to have in his "competence" the sum total of the rules relating form and meaning in a given language, but it seems clear that the comprehension of any utterance or discourse would entail the knowledge of whatever rules are actually applied in that utterance or discourse. Thus, the comprehension of a sentence like (2):

(2) The Fundalan added an are to his plot

would entail knowledge of such rules as the one whereby the suffix -an may imply "person originating from," the one indicating the possibility of the co-reference of Fundalan and his, the one whereby "are" is a noun denoting a unit of surface measure in the metric system, the rule specifying the meaning of the collocation "add" + "to," the rule specifying the meaning of "plot" as "a small piece of ground," and perhaps most important of all, the rules whereby the Fundalan,

added, and an are stand in subject-verb-object relationship, with the meaning of that relationship.

A major contribution of contemporary linguistic developments has been to bring out the richness of the semantic and grammatical rules underlying linguistic elements. The rather primitive conceptions of word meanings exemplified in certain kinds of psycholinguistic investigations, such as studies of word association and of "semantic differential" ratings, fail to do justice to this richness. We now know that even single words like "add," "are," and "plot" entail elaborate lexicogrammatical information with respect to the classes of experience to which they relate along with the kinds of grammatical constructions in which they can participate. Thus, in tracing the development of an individual's competence in a language one must take account not only of frequently studied morphological and syntactical phenomena such as pluralization and passivization, but also of the detailed lexicogrammatical knowledge about individual elements that participate in these phenomena. For example, in a recent study I found that whereas most 6th graders know the meaning of mill (as a noun) in the sentence "The children walked to the mill," relatively few comprehend mill (as a verb) in the sentence, "Before class, the children mill in the halls" (Carroll, 1970).

Having tried to give some specification of what we mean by "the meaning of a linguistic element," we may turn our attention to trying to characterize the "total meaning of an utterance," whatever the length of that utterance. Clearly, as Miller noted, the total meaning is not the sum total of the meaning of the words in the utterance. But now that we have defined "linguistic element" in such

a broad way as to include grammatical structures like the elements of phrase markers, it is tempting to conclude that the "total meaning" of an utterance is the sum total of the linguistic rules that have to be applied in the interpretation of the utterance, and that comprehension is therefore simply the application of these rules. Such a conclusion would correspond roughly to the proposal that has often been made that the comprehension of an utterance or discourse consists in the assignment of a "full structural description" to the message, if it is understood that such a structural description would have to include not only the ascription of a particular grammatical structure, but also the ascription of particular meanings to the constituents entering into that structure at various levels of analysis.

This solution does not seem completely satisfactory. One problem that arises is illustrated by the comprehender's task in assigning a meaning to "plot" in sentence (2). Suppose he knows that "plot" can mean either a "scheme, malicious plan" or "a small piece of ground." How does he know that in this sentence it means "small piece of ground"? That is, are there any linguistic rules that determine this? The kind of semantic theory developed by Katz and Fodor (1963) would probably answer that he knows it means "small piece of land" because both <u>are</u> and <u>plot</u> contain a common semantic feature of "surface area." In effect, the sentence signals that "the Fundalan added an area to his area," since a linguistic rule of interpretation would dictate that the meaning of "plot" should be selected in such a way as to accommodate its semantic features with those of other elements in the sentence. But such a rule may be gratuitous in the sense that it fails to honor the ability of the comprehender to

"make sense" of the sentence "on his own," thus without applying such a rule. And in fact a context for sentence (2) is (rather remotely, one must admit) conceivable wherein "plot" is to be interpreted as "malicious scheme." Moreover, the sentence is ambiguous in a number of other ways: Fundalan and his may or may not be co-referential, and Fundalan may or may not denote a "person of Fundala," since this word might denote some person of authority like a Nizam or a Mogul--it might even denote a nonhuman entity, as some sort of decree like the Magna Carta. In actual use of the sentence in a discourse, these ambiguities could only be resolved by information given in some wider context, either preceding or following the sentence. It is possible that discourse rules could be devised and invoked to specify how the disambiguation would take place, and if so, one might say that the correct comprehension of the total meaning of the sentence would involve the correct application not only of rules applying narrowly within the sentence but also of rules relating the sentence to its wider context. It remains to be seen, however, whether discourse rules having the kinds of potentialities envisaged here can in fact be formulated.

What does, at any rate, seem to be suggested by this consideration of ambiguity is that the "total meaning" of an utterance has to do with the relation of a sentence or discourse to its total context. If we widen the context beyond a mere "verbal" context, that is, to include the total situation in which the message occurs, its "total meaning" may entail the point-to-point relations between the elements encoded in the sentence and the things, attributes, events, and relations existing in some actual or fictional reality. Comprehension

of this "total meaning" would in this case imply awareness of these relationships. Thus, comprehension of sentence (2) would entail awareness of which Fundalan and which plot are referred to.

Suppose that sentence (2) occurs as the first sentence of a novel that is constructed in such a way that the full explanation of who or what the Fundalan was, and what was accomplished when an are was added to someone's plot, is disclosed only in the last chapter. If the "total meaning" of the sentence were held to be all these things, the gaining of that meaning is obviously a process that calls into play much more than a set of linguistic rules. This kind of "total meaning" would be best appreciated by a reader who returns to the first sentence after finishing the novel.

But what kind of comprehension could one expect when the reader reads the sentence for the first time? He could be expected at that point only to comprehend enough of it to get himself set to disambiguate the subsequent text at whatever pace the writer's design and the reader's patience would permit, and in this case we could say that comprehension entails the apprehension of just that amount of linguistic information that is "committed" to the sentence--information that could presumably be captured in a set of linguistic rules. Indeed, it might be part of the writer's design to leave the sentence ambiguous, allowing the reader to interpret it as he might. In such an interpretation, the predilection or disposition of the reader might be described probabilistically. For example, from past experience the reader would probably be more likely to infer the co-referentiality of Fundalan and his than the contrary. A joke-teller often deliberately leads a hearer into a misinterpretation of his opening narration

so that the "punch line," requiring another interpretation, will have its humorous effect.

This line of argument suggests that an "adequate" comprehension of a message at the time of its reception may be achieved by the comprehension of just that linguistic information that is "committed" to the message in terms of its own structure and in terms of whatever information has been disclosed by virtue of previous context. Some of this information may be of an ambiguous character, to be disambiguated by later information, provided that memory for the former is adequate. At a later time, comprehension of "total meaning" becomes more complete.

Our preliminary characterization of language comprehension may be summarized by stating that comprehension of a message is adequate or satisfactory to the extent that the language receiver apprehends, at least provisionally, whatever linguistic information is present in the message and is able to relate that information to whatever context is available at a given time. This implies that comprehension may be regarded as a process that contains at least two stages: (a) apprehension of linguistic information, and (b) relating that information to wider context.

There is a kind of paradox or inconsistency in this that I cannot see how to resolve at the moment: I have tried to distinguish "literal" or "plain-sense" comprehension from processes of inference, yet the relating of linguistic information to a wider context may indeed require processes of inference. For example, "adequate" comprehension of the second clause of a sentence such as:

(3) John isn't as tall as Mary, but Mary is shorter than he.

would entail the detection of the logical contradiction contained

there since the first clause provides the "wider context" to which the meaning of the second clause is to be related. Possibly one can resolve this contradiction by more closely identifying "literal" comprehension with the apprehension of linguistic information.

One may now ask what kind of comprehension can occur when messages are degraded in various ways. In natural situations, messages are often degraded by transmission failures, i.e., parts of the message do not reach the receiver. The concept of redundancy can and has been invoked to explain the fact that such a message can often be understood as well as, or nearly as well as, the original message; the redundancy may exist either purely among elements of linguistic information or between elements of linguistic information and some wider context. Nevertheless, redundancy is likely to involve probabilistic considerations in that a particular interpretation may become merely probable rather than certain.

Redundancy may also explain the fact that a subject in a psychological experiment such as the one conducted by Danks (1969) can claim to comprehend a scrambled, "ungrammatical" sentence such as (4):

(4) The helped nurse patient the.

even though interpretation may take somewhat longer, i.e., entail more processing of information, than it would if the sentence were unscrambled. The wider context contained in the subject's knowledge suggests, however, that the interpretation is more likely to be "The nurse helped the patient" than "The patient helped the nurse." Danks himself considers that the comprehension of deviant sentences of this type may be explained by

an appeal to "Ziffian" rules (Ziff, 1964) whereby the "simplest route" from the deviant sentence to a nondeviant sentence would be found, but I feel that something more than these rules must be invoked. For example, the Ziffian "inversion" rule would not explain why the subject is more likely to select one interpretation than another in the sentence cited, because there are two possible inversions.

In naturalistic contexts, one would be interested in the case of comprehension of "unclear" or "poor" writing. In general, it would seem inappropriate to expect the individual to comprehend more information than has been "committed" to the message itself, yet we know that readers (and hearers) are often able to "make sense out of" unclear messages by some as yet unexplicated inferential processes.

There is also the obverse case, that is, the case in which a language receiver fails to comprehend a message, or misinterprets it. According to our analysis of the comprehension process, this could occur at either one or both of the two stages, apprehension of linguistic information, and relating this information to wider context. That is, either the individual does not have the knowledge of the linguistic rules required to form a proper reading of a message, or he fails in the processing of that information, or both kinds of failure occur.

Even more generally, the kind of problem posed by this analysis is the explanation of what processes occur in what we have called "relating linguistic information to a wider context." The study of linguistic rules whereby language receivers gain certain types of information from messages is important, but equally important--and probably independent of purely linguistic study--is the study of how

the language user processes that information in order to assimilate or integrate it with his prior knowledge or cognitive structure.

The Testing of Comprehension

If the above analysis is correct, testing of comprehension involves consideration of the two conceptually separable stages of the comprehension process. That is, we would like to find out, in a given case, the extent to which the individual "correctly" apprehends the purely linguistic information that is "committed" to the message, and also the extent to which he "correctly" relates that information to some wider context.

There are several desiderata for tests of comprehension:

(1) Validity. An ideal test of comprehension should be valid in the sense that it reflects solely comprehension as defined here and not any other behavioral process such as memory, inference, guessing, or the like.

(2) Reliability. Ideally, a measure of comprehension should be reliable in the sense that it gives consistent outcomes on equivalent trials for a given individual.

(3) Generality. Ideally, a procedure for measuring comprehension should be applicable to (a) all types of verbal material, and (b) all classes of individuals. By "all types of verbal material," I have in mind variation in the quantity and complexity of the material--whether it be a single word, a single sentence, a paragraph, or a longer discourse, whether it be picturable or not, concrete or abstract, literary or technical in subject-matter, etc. By "all classes of individuals" I have in mind groups at different age levels, or with different degrees of competence in the language of the test.

(4) Convenience and practicality. The procedure should, ideally, be easy to prepare and easy to administer, and should yield outcomes that are easy to score or otherwise evaluate.

I have tried to develop a classification of procedures for testing comprehension on the basis of a survey of procedures followed either in psychometric devices or in experimental investigations. This proved to require a three-way classification in terms of (I) tasks, (II) types of measurements or observations taken, and (III) conditions of testing in terms of the temporal relations between presentation of the verbal stimulus and the taking of measurements or observations. Any given procedure can be classified as some combination of a particular task with a particular type of observational prodecure with some particular arrangement of the temporal relationships involved. While the classifications of tasks, types of measurements, and conditions of measurement do not completely exclude overlap, the framework has been useful in organizing the subsequent discussion.

I. Tasks

1. Subjective reports concerning:

(a) Comprehension vs. noncomprehension, degree of comprehension or comprehensibility

(b) Specific aspects of the message, e.g.:

(1) meaningfulness, analyticity, ambiguity, etc.

(2) grammaticality, "acceptability."

(3) "importance," "centrality," or "salience" of particular parts of the message.

2. Reports of truth or falsity, or of equivalence (in some sense) with another stimulus.
   (a) Analytic judgments
   (b) Verification with respect to another presentation
      (1) With respect to another message (to determine equivalence of meaning)
      (2) With respect to pictured referents
   (c) Verification with respect to the individual's knowlege base
3. Nonverbal response to the message: "following directions."
4. Supplying missing elements in a message
   (a) "Standard" cloze procedure (supplying missing words that have been deleted according to some rule)
   (b) "Progressive" cloze procedure (progressive adding of words, with feedback)
   (c) Sentence completions
   (d) Supplying order (as in an anagram or sentence rearrangement task)
5. Answering questions based on the message.
   (a) Completion-type items
   (b) Multiple-choice items
6. Recognition of messages, or elements thereof, on subsequent presentation
7. Reproduction of the message, in whole or in part, in original form or in some transformation
   (a) Verbatim reproduction
   (b) Paraphrase
   (c) Translation into another language or symbolism
   (d) The "probe latency" technique, e.g., reproduction of a given part of a message associated with a given cue
   (e) Eye-voice span (in reading)

II. Measurements or observations

1. Ratings or similar judgmental indices

2. "Correctness" of response with respect to some criterion

3. Time measurements

(a) Decision or response time

(b) Reading speed

(c) Learning time (or, number of trials)

4. Physiological responses

(a) Overt: emotional responses such as laughter, fear, etc.; eye movements

(b) Covert: electromyography, GSR, etc.

III. Conditions of testing

1. Responses elicited or observed simultaneously with message presentation

2. Responses elicited or observed immediately following message presentation

3. Responses elicited or observed after a delay.

(In 2. and 3. the original message, in whole or in part, may or may not be physically available during elicitation of the response.)

The following discussion of the various procedures for testing comprehension will be arranged according to the tasks required of the individual whose comprehension is being tested.

1. Subjective reports. Some remarks on subjective reports of comprehension have already been made. If the subject's "honesty" and attention can be assured, and particularly if accompanying measurements such as decision time can be taken, subjective reports would seem to be valid and highly useful measurements of comprehension. They have

been used only infrequently in psycholinguistic investigation, however (Danks, 1969), and the full potentialities of the method have not been explored. For example, the method might be used to explore what particular elements of a message cause difficulty in comprehension, e.g., particular words, grammatical constructions, clauses, etc. By varying the nature of the message, as Danks did, it is possible to relate subjective ratings and decision times to message characteristics such as grammaticality, ambiguity, grammatical complexity, vocabulary difficulty, etc. Kershner (1964) measured reading times for passages of different levels of difficulty, both before and after the subject learned that he was going to be required to answer questions on a passage. The amount of time taken by the subject to read a passage may be thought of as reflecting the judgment of the subject as to whether he understands it.

While subjective reports could easily yield false positive results when the individual believes himself to comprehend, but actually does not, it is unlikely that they would yield false negative results unless the individual is malingering. The presence of false positive results could be detected by use of certain other techniques, such as asking questions. If subjective reports of comprehension are taken simultaneously with, or immediately after, presentation of the message, memory factors will have little or no influence. The extent to which subjective reports of comprehension will reflect inferential processes would probably depend upon the degree to which the message requires the operation of such processes.

Unlike the remainder of the techniques, subjective reports of comprehension cannot be used in an adversary testing situation; the subject would be too likely to claim comprehension falsely.

2. Reports of truth or falsity, or of equivalence (in some sense) with another presentation. When verification of a message can be based either on the analyticity of the message or upon, say, a pictured referent, this technique has much to recommend it as a measurement of pure comprehension, because (if the subject is honest and attentive), a correct response is directly dependent upon comprehension. The technique has many of the features of the subjective report; in fact, it is a kind of subjective report of comprehension. On the other hand, when verification is against the knowledge base of the individual (e.g., "The capital of South Africa is Johannesburg: True or False?") it is more likely to measure that knowledge base than the presence of comprehension.

Because of the simplicity of the binary judgments required, the measurements may suffer from unreliability and therefore may have to be buttressed by additional measurements (replication, use of feedback and correction, and the like). Wason (1961) used this method in an experiment on the comprehension of negation; he measured the latency of judgments of the truth or falsity of analytic sentences like "88 is not an even number" and pooled the results over samples of such sentences. Nevertheless, Ss made relatively few errors. Extensive use of picture verification procedures has been made by Slobin (1966) and Gough (1965, 1966), with precautions similar to those taken by Wason. Gough experimentally varied the time relations between presentation of the verbal message and the picture.

An extension of this technique, particularly appropriate for listening comprehension, but also useful for reading comprehension, is to present a sentence and require S to choose which of several

pictures best represents its meaning. Alternative choices can be designed to require S to make fine discriminations among linguistic elements. Its major disadvantages are its inconvenience (the difficulty of drawing satisfactory pictures) and the fact that there is probably a limit to what can be presented in pictorial form.

Another variant of this general technique would be to have S evaluate whether a given message is equivalent in some respect (e.g., meaning) to another message. A simple and common form of this procedure is to be found in vocabulary tests, where S is required to select a word similar in meaning to a key word. As applied to larger units such as sentences, the technique has received little use (unless one considers that certain types of multiple-choice comprehension tests are a variant of this technique).

3. Nonverbal responses to a message: following directions. Tests of the subject's ability to follow verbal directions by carrying out some performance have appeared in intelligence tests ever since the construction of the Army Alpha test in World War I, but have rarely been used in experimental studies of comprehension, despite the fact that such tests could be highly valid, reliable and convenient measurements in many circumstances. Jones (1966) had children perform a cancellation task under instructions such as "Mark all the numbers [in a display] except 2, 5, 8." Shipley, Smith, and Gleitman (1969) tested children's comprehension by having them execute commands. Another variant of the technique has been effectively employed by Carol Chomsky (1969).

To insure validity, however, the task must be one that is not likely to be performed correctly unless S has understood the instructions. The procedure has the disadvantage that it may be applicable only to a certain

limited set of verbal materials, and it may be subject to the influence of memory factors in that S may comprehend the instructions but forget them before he begins to perform the task.

4. Supplying missing elements in messages. The most typical and popular example of this technique is the "cloze" procedure introduced (or reintroduced) by Taylor (1953) initially as a measure of "readability" (the difficulty of a text). The procedure involves taking a passage of text and deleting words in it by some rule, e.g., every 5th word, every other noun, or every other "function" word. A subject is then presented with the passage and asked to guess the missing words. Usually the passage is presented in written form, in which case the missing words are indicated by blanks of a standard size, but techniques are also available for presenting the passage in auditory form (Peisach, 1965). The procedure has gained considerable acceptance as a measure of the individual's degree of comprehension of a given text (Bormuth, 1968; Greene, 1965; Taylor, 1957). Such measures are found to have substantial or even high correlations with more conventional tests of reading comprehension.

The validity of the "cloze" technique in measuring an individual's comprehension of a given text is open to some question. Weaver and Kingston (1963) performed a factor-analytic study that suggested that scores are affected by a special aptitude or ability for utilizing redundancy in a passage, and supplying missing elements, independent of verbal ability. Coleman and Miller (1968) tried to use the technique in measuring knowledge gained from prior inspection of the unmutilated passage but found that the scores were hardly higher, on the average, than those of Ss who had not been presented with the unmutilated passage. It would seem that cloze scores are dependent chiefly on what might be called the "local redundancy" of a passage, i.e., the extent to which

linguistic cues in the immediate environment (generally, in the same sentence) of a missing word tend to supply it. Rankin (1958) found that cloze scores based on deletions of nouns and verbs seem to measure something other than what is measured by scores based on deletions of function words. There is no clear evidence that cloze scores can measure the ability to comprehend or learn the major ideas or concepts that run through a discourse. It is even possible to secure cloze scores on the basis of meaningless material so long as grammatical cues are present; thus, cloze scores are probably more dependent on detection of grammatical than of semantic cues. On the whole, the cloze technique in its usual form is too crude to permit measuring the degree to which the individual comprehends particular lexical or grammatical cues, or possesses a knowledge of specified linguistic rules. It probably depends to a considerable extent on inferential processes.

The "progressive cloze" technique requires the subject to guess each successive word of a passage. Rubenstein and Aborn (1958) allowed only one guess per word (but gave the correct word after each guess) and measured the difficulty of passages in terms of the percentage of words correctly guessed by a group of subjects. These scores were highly correlated with readability and learning scores obtained from other subjects. This illustrates use of the technique in scaling passage difficulty. Coleman and Miller (1968), however, used it in measuring an individual's ability to learn from a passage. Essentially, their procedure had the subject take two trials with the same passage. The gain in the percentage of correct guesses on the second trial was considered a measure of information gained through exposure on the first trial. Because of the interval between a guess on the first trial and a guess on the second trial their technique necessarily involves a memory factor and is thus not a pure measure of comprehension.

There are certain other forms of comprehension tests that require the supplying of missing elements from context and that are more highly focussed on testing the comprehension of particular types of cues. For example, a sentence may be given in which the supplying of the one missing word would be contingent (at least partly) on the detection of a particular grammatical or lexical cue. Sentence completion tests have been used in studies of grammatical ambiguity: the type of completion supplied by the subject indicates the particular interpretation he makes for an ambiguous expression (MacKay, 1966). When sentences are presented in a scrambled arrangement, the missing elements consist of the cues of word order that are present in normal text (Oléron, 1961); in reconstructing the text, the subject has to supply these elements from other types of cues.

5. Answering questions based on the message. One finis on nearly all standardized reading or listening comprehension tests the device of presenting a paragraph to read or listen to, with one or more questions to be answered over the content of the paragraph. Ordinarily, on reading tests this paragraph is available to the subject as he answers the questions; there is little control of the subject's strategy, and some subjects believe they will do better if they read the questions before they inspect the paragraph. In listening tests, the questions are usually given after the presentation of the message and the subject has to depend on memory. Since the object is generally to measure comprehension ability, the selection of items is controlled by statistics concerning whether the correct answers on the individual items are correlated with scores on the test as a whole or with some external criterion such as scholastic success. Scores on these tests are often highly correlated with measures of general verbal ability.

There is evidence that depending on the form and content of the questions, different kinds of reading or listening "skills" can be measured (Bateman, Frandsen, & Dedmon, 1964; Davis, 1968).

It is too often the case that the questions on reading and listening comprehension tests are not controlled for the ability of the subject to answer them above a chance level even if they are not exposed to the texts on which the questions are based. Often the questions can be answered on the basis of the subject's prior knowledge or on the basis of various incidental cues in the questions themselves. Sometimes the questions present difficulties that are extraneous to the comprehension of the text. A technique for controlling such factors has been presented by Marks and Noll (1967).

The construction of items for comprehension tests has traditionally been viewed as a matter requiring much ingenuity, creativity, and even artistry on the part of the item-writer. Bormuth (1970) has severely (and perhaps unjustly) criticized traditional test-construction procedures for their unsystematic, "unscientific" nature and suggests that a science of item-construction can be developed by using principles of transformational grammar. It remains to be seen whether such a suggestion can in fact lead to measurements of all the aspects of comprehension and learning that one might want to measure, but Bormuth's techniques have much promise for testing the individual's ability to apprehend the information provided by purely linguistic cues.

6. Recognition of messages, or elements thereof, on subsequent presentation. The recognition technique has been a traditional method of measuring learning and memory. The subject is presented with an array of material that he is asked to inspect or learn, after which (either immediately or after a delay) he is given elements of the

original array together with new or modified elements and asked to indicate which elements are "old" and which are new. For example, Shepard (1967) asked college-age students to inspect, one by one, 612 short, unrelated sentences, after which they had to identify, in a series of 68 test pairs, which member of each pair had occurred in the previous series; they were 89% accurate in doing so (chance success being 50%). Since the sentences were all easily comprehensible on first presentation, the results undoubtedly reflect memory rather than comprehension processes.

Nevertheless, the recognition technique has been used by several investigators to examine detailed processes of comprehension. Clifton, Kurcz, and Jenkins (1965), and Clifton and Odom (1966) used a recognition task to index the grammatical similarity of sentences; after presentation of a series of sentences, these same sentences together with grammatical variants of them (involving negative, passive, and question transformations) were presented and the subject was asked to press a telegraph key whenever he thought he recognized one of the "old" sentences. Fillenbaum (1970), however, has shown that this technique was inadequate to capture subtle semantic differences among sentences. Lee (1965), Fillenbaum (1966), Newman and Saltz (1960), and Sachs (1967a, 1967b) have used the recognition task to find out the extent to which subjects remember the <u>verbatim</u> forms of words or sentences as opposed to their meanings. The evidence indicates, in general, that verbatim forms are remembered only for a relatively short time, if at all, whereas meanings are remembered much longer.

Another application of the recognition technique is the "chunked comprehension" test developed by Carver (1970). Carver presents a passage for reading, typically four or five paragraphs long. This is

then immediately followed by a multiple-choice test that the examinee must complete without referring to the original passage. In each item of the multiple-choice test, each alternative consists of a "chunk" of the original--a clause, a phrase, or sometimes a single word; one "chunk," however, is changed in meaning by the substitution of a different word or phrase. The subject has to indicate which alternative does not convey the original meaning. An example will illustrate the technique. The first paragraph of one of Carver's selections is as follows:

> Voter apathy is almost a cliché in discussions of American politics. Yet, only a cursory look at voting and registration restrictions shows that many would-be voters do not cast ballots because they are prevented from doing so.

The test items covering this part of the selection are as follows:

1. (A) Voter apathy
   (B) is almost a cliché
   (C) in discussions
   (D) of American politics.
   (E) A recent poll directed
2. (A) at voting
   (B) and registration restrictions
   (C) shows that
   (D) many would-be voters
   (E) seldom protest or demonstrate
3. (A) because they are prevented
   (B) from doing so.
   (C)
   (D)
   (E) [The remaining alternatives cover the beginning of the next paragraph in the selection.]

The changed alternatives are constructed and item-analyzed in such a way that individuals who have not read the original passage are unable to score much above chance; doubtless this process requires much ingenuity and experimentation.

By definition, the recognition technique reflects memory processes. Even if comprehension processes are involved, it is difficult to separate their effects from those of memory processes. Thus, Carver's

"chunked comprehension" test cannot be regarded as a measure of comprehension as such; in fact, the manual for the published version of the test (Darby & Carver, 1970) states that it is designed to test "memory storage" for verbal content. It is a test of comprehension only to the extent that memory processes may be assumed to be solely a function of degree of comprehension, at least in the test situation. Some support for such an assumption can be found in Underwood's (1964) suggestion that amount of retention, when temporal factors are controlled, is chiefly a function of degree of original learning. Even so, this would imply that the recognition technique can be used to index comprehension only when there is precise control of temporal factors.

7. Reproduction of the message, in whole or in part, in original form or in some transformation. An extraordinary variety of techniques for testing or investigating language comprehension or verbal learning involve tasks requiring reproduction of a message in some form. Depending on the nature of the task and the conditions of testing, memory processes may be involved, and thus, as in the case of the recognition task just discussed, the respective roles of comprehension and memory processes may be difficult to isolate.

For example, verbatim recall of single sentences immediately after visual or auditory presentation may depend either on pure memory span or upon comprehension, or some combination thereof. There is no systematic body of information about memory span for verbal material. Miller (1956) reports data from Hayes that indicates that the memory span for unrelated words is above 5 for mature speakers. As soon as there is any degree of semantic or syntactic organization in a series of words presented for immediate recall, the number of words

that can be recalled increases beyond the span for unrelated, meaningless materials (Marks & Jack, 1952). This is not to say, however, that short-term memory factors cease to operate. Memory span for well-formed sentences has been considered an index of mental age (Terman, 1916, pp. 37-39). It has also been used in the study of the development of linguistic competence in young children (e.g., Slobin & Welsh, 1968).

The experimental study of verbatim reproduction of longer passages (Clark, 1940; Henderson, 1903; Lyon, 1917) has generally depended on a scoring procedure known as the "method of retained members." The stimulus passage is divided into a number of phrasal units of approximately equal size; the subject's response is then scored in terms of the number of these units that are reproduced. Levitt (1956) showed that different investigators are likely to make different divisions of a passage and these differences are likely to be reflected in recall scores. There seems to have been no application of strictly linguistic procedures to determine what units should be scored. King (1960, 1961) and his collaborators (King & Russell, 1966; King & Yu, 1962) have reported a series of studies showing that when judges are asked to scale written recalls for excellence, two factors influence their judgments: a "quantitative" factor having to do with the amount of recall (number of words, and the like), and an "organization" factor having to do with the quality and organization of the semantic content. This result implies, incidentally, that judges differ in the extent to which they are influenced by these factors.

One of the more perceptive studies of verbatim recall that I have found was by Gomulicki (1956), who presented his subjects with 37 prose passages, from 13 to 95 words in length. He studied the reproduction of each word, judging it as either "adequate" or "inadequate." Over

the whole set of reproductions, 55.5% words were reproduced verbatim, 32.7% were omitted, 11.8% were changed, and 6.2% were added words or ideas. The frequency with which a given element was "adequately" represented was regarded as a measure of its "mnemic value." Mnemic value was then studied as a function of semantic content (action vs. description) and grammatical function. Recall was regarded as an "abstractive process." The best remembered materials described actor-action-effect sequences; there was even a tendency for Ss to turn descriptive passages into "quasi-narratives."

Immediate verbatim recall of verbal materials has been used to study many aspects of language behavior and learning: basic processes in recall (Bartlett, 1932; Paul, 1959); the effect of "order of approximation to English" (Miller & Selfridge, 1950; Tulving & Patkau, 1962); the effect of syntax and other grammatical factors (Miller, 1962; Slobin & Welsh, 1968); the effect of instructions as to what is to be recalled (King & Russell, 1966); the effect of associational factors (Rosenberg, 1968); and oral vs. printed stimuli (King & Madill, 1968).

Space does not permit discussion of the many variants of the recall task: delayed verbatim recall (Slamecka, 1959); recall after interpolated material (Savin & Perchonock, 1965); time for verbatim learning to a criterion (Follettie & Wesemann, 1967; Rubenstein & Aborn, 1958); paired-associate learning in which sentences are the responses (Martin & Jones, 1965); serial learning of sentences (Epstein, 1962); etc. Although the effects of various message characteristics (meaningfulness, grammatical structure, etc.) on the recalls can be studied by appropriate experimental controls, it remains difficult to differentiate comprehension, storage, and retrieval processes.

There are several special variants of the message-reproduction task that deserve consideration. One is the paraphrasing task, i.e., reproducing the message in the subject's "own words." Generally it is required that this task be performed without the subject's being able to refer to the original message, but if memory processes are to be excluded, this need not necessarily be the case. If paraphrases can be objectively and validly scored, this task may be a useful technique for measuring comprehension. The catch is that it may be very difficult to score paraphrases for conformity of content to the original, as was noted for example by Downey and Hakes (1968). Moreover, telling the subject to use his "own words" may place an extra burden on him when he interprets this as meaning that he cannot use the words of the original message. And, of course, it is possible for paraphrases to be nothing more than grammatical transformations performed without full comprehension of semantic content.

The writer (Carroll, 1970) recently used a paraphrase task to study children's comprehension of single words used in unusual grammatical functions; the words in question were placed in imaginary "headlines" such as WHEN YOU ARE LOST, SOMEONE WILL PAGE YOUR MOTHER. High reliability in scoring the responses was achieved, but it was probably the case that some unsuccessful responses reflected simple inability to create a paraphrase even though the respondent actually comprehended the sense of the message; this would be an example of a false negative outcome.

Translating a message into another language is a traditional method of assessing comprehension in foreign-language learning, as where an English-speaking student is required to translate a French sentence or paragraph into English. Obviously, this method cannot

be generally used in testing native-language comprehension, and even in foreign language instruction there is the problem of attaining adequate scorer reliability, not to mention the problem of defining what a truly adequate translation is.

The translation of verbal messages into mathematical or logical symbolism might appear to be an analogous possibility. I have in mind the kind of comprehension required, for example, in order to state an algebraic formula for the solution of a verbally-stated mathematical problem. I have not looked into the research literature concerning this problem, as there are obvious drawbacks to the generality of the procedure (the respondent's knowledge of the mathematical or logical symbolism involved would be a factor, certainly).

The "eye-voice span" in reading a text has been used by several investigators (e.g., Levin & Kaplan, 1966; Schlesinger, 1966) as an index of comprehension processes. It can be regarded as a variant of the reproduction task, in that the subject is required to reproduce that part of a printed message that is within his span of perception but not yet read aloud, in an oral reading task in which the subject's viewing of the stimulus is suddenly terminated at a particular moment. Presumably, the eye-voice span reflects the additional information processing that the subject is performing on material ahead of what he is reading aloud at that moment. While it may represent the operation of sentence-comprehension processes, it may also reflect certain inferential and guessing processes similar to those tapped in the "cloze" technique.

* * * * * *

This brief survey of techniques that have been used to test language comprehension points up the fact that there is no one technique

that universally gives valid and reliable information. It is seldom the case that success or failure in any of these tests can unequivocally be traced to success or failure in language comprehension since there are other factors of guessing, inference, memory, reliance on prior knowledge, etc., that are operating. The influences of these other factors must be controlled as fully as possible by variation of message characteristics, control of temporal factors, and instructions to the subject.

In this discussion, not much has been said about the capability of the techniques to distinguish the two processes earlier identified as inherent in comprehension: apprehension of linguistic information, and relating that information to a wider context. Psycholinguistic investigations have, for the most part, ignored this problem. Little context is offered when single sentences are presented, and when the comprehension of longer discourse has been studied, there has been little attempt to explicate contextual elements or to vary them experimentally. Whether such an approach would be useful remains to be seen.

References

Bartlett, F. C. Remembering. Cambridge, England: Cambridge University Press, 1932.

Bateman, D., Frandsen, K., and Dedmon, D. Dimensions of "lecture comprehension": A factor analysis of listening test items. Journal of Communication, 1964, 14, 183-189.

Blommers, P. J., and Lindquist, E. F. Rate of comprehension of reading: Its measurement and its relation to comprehension. Journal of Educational Psychology, 1944, 35, 449-473.

Bormuth, J. R. Cloze test readability: Criterion reference scores. Journal of Educational Measurement, 1968, 5, 189-196.

Bormuth, J. R. On the theory of achievement test items. With an appendix by P. Menzel: On the linguistic bases of the theory of writing items. Chicago: University of Chicago Press, 1970.

Bormuth, J. R., Manning, J., Carr, J., and Pearson, D. Children's comprehension of between- and within-sentence syntactic structures. Journal of Educational Psychology, 1970, 61, 349-357.

Brown, J. I., and Carlson, G. R. Brown-Carlsen Listening Comprehension Test. New York: Harcourt, Brace and World, 1953.

Carroll, J. B. A factor analysis of verbal abilities. Psychometrika, 1941, 6, 279-307.

Carroll, J. B. From comprehension to inference. In M. P. Douglas (Ed.), Thirty-Third Yearbook, Claremont Reading Conference. Claremont, Calif.: Claremont Graduate School, 1969. Pp. 39-44.

Carroll, J. B. Comprehension by 3rd, 6th, and 9th graders of words having multiple grammatical functions. Final Report, Educational Testing Service, Princeton, N. J., December 1970. Project No. 9-0439, Grant No. OEG-2-9-400439-1059, U. S. Office of Education.

Carver, R. P. Analysis of "chunked" test items as measures of reading and listening comprehension. Journal of Educational Measurement, 1970, 7, 141-150.

Chomsky, C. The acquisition of syntax in children from 5 to 10. Cambridge, Mass.: M. I. T. Press, 1969.

Clark, H. H. Linguistic processes in deductive reasoning. Psychological Review, 1969, 76, 387-404.

Clark, K. B. Some factors influencing the remembering of prose materials. Archives of Psychology, New York, 1940, No. 253.

Clifton, C., Jr., Kurcz, I., and Jenkins, J. J. Grammatical relations as determinants of sentence similarity. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 112-117.

Clifton, C., Jr., and Odom, P. Similarity relations among certain English sentence constructions. Psychological Monographs: General and Applied, 1966, 80(5).

Coleman, E. B., and Miller, G. R. A measure of information gained during prose learning. Reading Research Quarterly, 1968, 3, 369-386.

Danks, J. H. Some factors involved in the comprehension of deviant English sentences. (Doctoral dissertation, Princeton University), Ann Arbor, Mich.: University Microfilms, 1969. No. 69-2735.

Darby, C. A., Jr., and Carver, R. P. Manual for the Carver-Darby Chunked Reading Test. Silver Spring, Md.: American Institutes for Research, 1970.

Davis, F. B. Fundamental factors of comprehension in reading. Unpublished doctoral dissertation, Harvard University, 1941.

Davis, F. B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.

Davis, F. B. Research in comprehension in reading. Reading Research Quarterly, 1968, 3, 499-545.

Downey, R. G., and Hakes, D. T. Some psychological effects of violating linguistic rules. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 158-161.

Edfeldt, Å. W. Silent speech and silent reading. Chicago: University of Chicago Press, 1960.

Educational Testing Service. Sequential Tests of Educational Progress: Listening. Princeton, N. J.: ETS, 1956-1959.

Epstein, W. A further study of the influence of syntactic structure on learning. American Journal of Psychology, 1962, 75, 121-126.

Epstein, W. Recall of word lists following learning of sentences and of anomalous and random strings. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 20-25.

Fillenbaum, S. Memory for gist: Some relevant variables. Language and Speech, 1966, 9, 217-227.

Fillenbaum, S. On the use of memorial techniques to assess syntactic structures. Psychological Bulletin, 1970, 73, 231-237.

Follettie, J. F., and Wesemann, A. F. Effects of grammatical factors and amount of material on memorizing paragraphs, sentences and word lists. Human Resources Research Organization, No. 67-9, 1967. (Document AD 656 454, Clearinghouse for Federal Scientific and Technical Information.)

Freshley, D. L., and Anderson, H. E., Jr. A factor analytic study of listening. Paper presented at the meeting of the American Educational Research Association, Chicago, February 1968.

Gomulicki, B. R. Recall as an abstractive process. Acta Psychologica, 1956, 12, 77-94.

Gough, P. B. Grammatical transformations and speed of understanding. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 107-111.

Gough, P. B. The verification of sentences: The effects of delay of evidence and sentence length. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 492-496.

Greene, F. P. A modified cloze procedure for assessing adult reading comprehension. (Doctoral dissertation, University of Michigan), Ann Arbor, Mich.: University Microfilms, 1965. No. 65-5308.

Henderson, E. M. A study of memory for connected trains of thought. Psychological Monographs, 1903, 5 (6, Whole No. 23).

Jones, S. The effect of a negative qualifier in an instruction. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 497-501.

Katz, J. J., and Fodor, J. A. The structure of a semantic theory. Language, 1963, 39, 170-210.

Kelly, C. M. An investigation of the construct validity of two commercially published listening tests. Speech Monographs, 1965, 32, 139-143.

Kershner, A. M. Speed of reading in an adult population under differential conditions. Journal of Applied Psychology, 1964, 48, 25-28.

King, D. J. On the accuracy of written recall: A scaling and factor analysis study. Psychological Record, 1960, 10, 113-122.

King, D. J. Scaling the accuracy of recall of stories in the absence of objective criteria. Psychological Record, 1961, 11, 87-90.

King, D. J., and Madill, J. Complex methods of presentation, internal consistency of learning material, and accuracy of written recall. Psychological Reports, 1968, 22, 777-782

King, D. J., and Russell, G. W. A comparison of rote and meaningful learning of connected meaningful material. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 478-483.

King, D. J., and Yu, K. C. The effect of reducing the variability of length of written recalls on the rank order scale values of the recalls. Psychological Record, 1962, 12, 39-44.

Lee, W. Supra-paragraph prose structure: Its specification, perception, and effects on learning. Psychological Reports, 1965, 17, 135-144.

Levin, H., and Kaplan, E. Studies of oral reading: X. The eye-voice span for active and passive sentences. Unpublished manuscript, Cornell University, September 1966.

Levitt, E. E. A methodological study of the preparation of connected verbal stimuli for quantitative memory experiments. Journal of Experimental Psychology, 1956, 52, 33-38.

Lyon, D. O. Memory and the learning process. Baltimore: Warwick and York, 1917.

Lyons, J., and Wales, R. J. (Eds.) Psycholinguistics papers: The proceedings of the 1966 Edinburgh conference. Edinburgh: Edinburgh University Press, 1966.

MacKay, D. G. To end ambiguous sentences. Perception and Psychophysics, 1966, 1, 426-436.

Marks, E., and Noll, G. A. Procedures and criteria for evaluating reading and listening comprehension tests. Educational and Psychological Measurement, 1967, 27, 335-348.

Marks, M. R., and Jack, O. Verbal context and memory span for meaningful material. American Journal of Psychology, 1952, 65, 298-300.

Martin, J. G., and Jones, R. L. Size and structure of grammatical units in paired-associate learning at two age levels. Journal of Experimental Psychology, 1965, 70, 407-411.

McGuigan, F. J., Keller, B., and Stanton, E. Covert language responses during silent reading. Journal of Educational Psychology, 1964, 55, 339-343.

Miller, G. A. The magical number seven plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.

Miller, G. A. Some psychological studies of grammar. American Psychologist, 1962, 17, 748-762.

Miller, G. A. Some preliminaries to psycholinguistics. American Psychologist, 1965, 20, 15-20.

Miller, G. A., and Selfridge, J. A. Verbal context and the recall of meaningful material. American Journal of Psychology, 1950, 63, 176-185.

Newman, S. E., and Saltz, E. Effects of contextual cues on learning from connected discourse. American Journal of Psychology, 1960, 73, 587-592.

Oléron, G. Compréhension d'un texte et liaisons entre les mots. Année Psychologique, 1961, 61, 377-395.

Paul, I. H. Studies in remembering: The reproduction of connected and extended verbal material. In G. S. Klein (Ed.), Psychological Issues. New York: International Universities Press, 1959. Part 2.

Peisach, E. C. Children's comprehension of teacher and peer speech. Child Development, 1965, 36, 467-480.

Rankin, E. F., Jr. An evaluation of cloze procedure as a technique for measuring reading comprehension. (Doctoral dissertation, University of Michigan) Ann Arbor, Mich.: University Microfilms, 1958. No. 58-3722.

Rosenberg, S. Language habits and the recall of connected discourse. In E. M. Zale (Ed.), Proceedings of the Conference on Language and Language Behavior. New York: Appleton-Century-Crofts, 1968. Pp. 63-74.

Rubenstein, H., and Aborn, M. Learning, prediction, and readability. Journal of Applied Psychology, 1958, 42, 28-32.

Sachs, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 1967, 2, 437-442. (a)

Sachs, J. S. Recognition of semantic, syntactic and lexical changes in sentences. Paper presented at the meeting of the Psychonomic Society, Chicago, Illinois, October 1967. (b)

Savin, H. B., and Perchonock, E. Grammatical structure and the immediate recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 348-353.

Schlesinger, I. M. The influence of sentence structure on the reading process. Technical Report No. 24, October 1966, The Hebrew University, Jerusalem, Contract No. N 62558-4695 NR 049-130, U. S. Office of Naval Research. (Document AD 653 263, Clearinghouse for Federal Scientific and Technical Information.)

Shepard, R. N. Recognition memory for words, sentences, and pictures. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 156-163.

Shipley, E. F., Smith, C. S., and Gleitman, L. R. A study in the acquisition of language. Language, 1969, 45, 322-342.

Slamecka, N. J. Studies of retention of connected discourse. American Journal of Psychology, 1959, 72, 409-416.

Slobin, D. I. Grammatical transformations and sentence comprehension in childhood and adulthood. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 219-227.

Slobin, D. I., and Welsh, C. A. Elicited imitations as a research tool in developmental psycholinguistics. Working paper No. 10, May 1968, Berkeley, Calif.: Language Behavior Research Laboratory, University of California.

Spearritt, D. Listening comprehension--a factorial analysis. Melbourne: Australian Council for Educational Research, 1962.

Taylor, W. L. "Cloze procedure": A new tool for measuring readability. Journalism Quarterly, 1953, 30, 415-433.

Taylor, W. L. Cloze readability scores as indices of individual differences in comprehension and aptitude. Journal of Applied Psychology, 1957, 41, 12-26.

Terman, L. M. The measurement of intelligence. Boston: Houghton Mifflin, 1916.

Thurstone, L. L. Note on a reanalysis of Davis' reading tests. Psychometrika, 1946, 11, 249-255.

Tulving, E., and Patkau, J. F. Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. Canadian Journal of Psychology, 1962, 16, 83-95.

Underwood, B. J. Degree of learning and the measurement of forgetting. Journal of Verbal Learning and Verbal Behavior, 1964, 3, 112-129.

Wason, P. C. Response to affirmative and negative binary statements. British Journal of Psychology, 1961, 52, 133-142.

Weaver, W. W., and Kingston, A. J. A factor analysis of the cloze procedure and other measures of reading and language ability. Journal of Communication, 1963, 13, 252-261.

Ziff, P. On understanding "understanding utterances." In J. A. Fodor and J. J. Katz (Eds.), The structure of language. Englewood Cliffs, N. J.: Prentice-Hall, 1964. Pp. 390-399.